Studies on High-Resolution Remote Sensing Sugarcane Field Extraction based on Deep Learning

Studies on High-Resolution Remote Sensing Sugarcane Field Extraction based on Deep Learning

Ming Zhu[1*,2], Maohua Yao[2], Yuqing He[2], Yongning He[2] and Bo Wu[2]

[1]Institute of Geoscience and Resources, China University of Geosciences, Beijing, 100083, China

[2]Geographic Information Center of Guangxi, Nanning, 530023 China

[*]Corresponding author's e-mail: 25568571@qq.com

Abstract. Sugarcane is one of the most important economic crops in Guangxi. For a long time, the sugarcane cultivated areas were estimated via sampling data statistics, while effective and accurate dynamic monitoring data keep absent. High spatial resolution is one of the advantages of high-resolution remote sensing images, through which the texture of sugarcane fields is found clear and unique; however, effective and accurate methods are lacking extracting them automatically in the past. In this paper, a novel deep learning method for sugarcane field extraction from high-resolution remote sensing images is proposed based on DeepLab V3+. It consists of blocks for multi-temporal remote sensing images fusion, which increases the ability of DCNN temporal factors processing. The experiment shows 94.32% extraction accuracy of sugarcane field. Also, its processing speed is superior to the traditional object-oriented extraction method, which solves the problems of low extraction accuracy and slow processing speed using traditional methods.

1. Introduction

Sugarcane is one of the most important agricultural products in Guangxi and its annual output ranked first in China every year since the 90s. In Guangxi, sugarcane planting is very important to economic development; however, the sugarcane cultivated areas were estimated via sampling data statistics for a long time when effective and accurate dynamic monitoring data keeps absent. In recent years, with the maturity of China's domestic high-resolution satellites, the spatial resolution and temporal resolution of remote sensing images have been continuously improved [1]. Sources variety of remote sensing satellites promises remote sensing applications a solid foundation of data acquisition. Sugarcane field can be dynamically monitored and accurately extracted from high-resolution remote sensing images through interpretation; whereas, its extraction mainly relies on artificial visual interpretation in practices. Although technologies based on artificial neural network, support vector machine, genetic algorithm and object-oriented have developed rapidly and made considerable achievements in application domains [2], these methods require manual designations on image features and interpretation rules, whose design cycle is long, potentials of algorithm improvement are limited, and whose accuracy and efficiency fail to meet the needs of many applications. The situation that remote sensing applications rely heavily on human labor, experience and skills, has severely restricted application performance of remote sensing[3].

Deep learning is an important branch of machine learning. It provides an end-to-end machine

learning model based on the Deep Convolutional Neural Network (DCNN), which can automatically extract image features without manual feature extraction[4]. Compared with traditional technologies, deep learning is completely data-driven, and it is able to obtain the best feature extraction method automatically through learning[5]. For this reason, deep learning method has been rapidly introduced into remote sensing research, and it became one of the research hotspots in remote sensing[6][7][8].

2. Data and methodology

2.1 Data Source
The GF-2 and BJ-2 high-resolution remote sensing images of Wuming, Guangxi were captured respectively on February, May, July and August 2017 with 0.8m spatial resolution and multispectral bands (red, green, blue and near infrared). Vector data include surface slope, National Geography Condition Monitoring, and double-high (high sugar and high yield) sugarcane base boundary. Slope data is used to improve the accuracy of deep learning extraction. The rest datasets are used to form training samples database and test extraction accuracy.

2.2 Methodology
Sugarcane cultivated area extraction and calculation technology belongs to the domain of semantic image segmentation. In this study, deep neural network is improved by DeepLab V3+. DeepLab[9] is a high-performance image segmentation network proposed by Liang-Chieh Chen et al. The latest version DeepLab V3+ utilizes pre-trained ResNet on ImagNet as the main feature extraction network. It also explores the Xception model and applies depth wise separable convolution to both Atrous Spatial Pyramid Pooling (ASPP)[10] and decoder modules, resulting in a faster and stronger encoder-decoder network. Through these improvements, the segmentation accuracy of DeepLab V3+ is maintained while Dense CRF is discarded, which is an imported part in the old version[11].
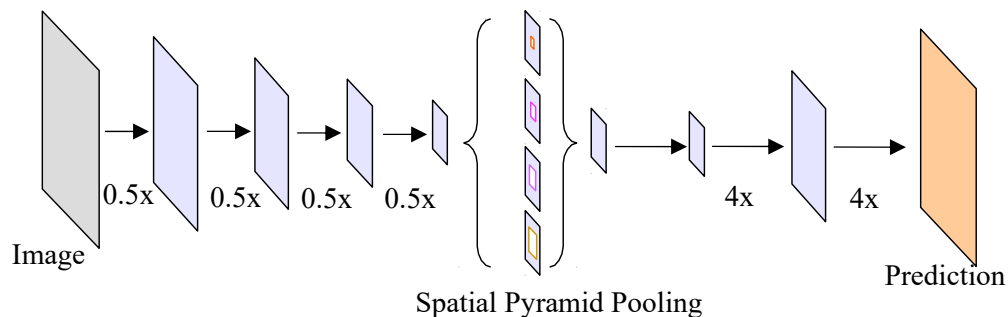

Figure 1. The Atrous Conv Structure of DeepLab V3+

The essence of high-resolution remote sensing sugarcane field extraction based on deep learning is to use Deep Convolution Neural Network (DCNN) to automatically extract the image features of sugarcane through a lot of sample trainings, and then use DCNN during the process of field extraction. Therefore, in the perspective of texture, higher homogeneity among a certain land cover type and bigger discrepancy towards the others in remote sensing images are propitious to conduct better recognition performance in deep neural network. However, most land in Guangxi is densely vegetated, and the textures of sugarcane fields appear differently along with sugarcanes' growth. The texture of sugarcane fields in the sowing and seedling period (March to April) is similar to that of bare land. During growing period (May to June), it is similar to mixture of crops and cultivated land. In the mature period (July-December), the texture presents vegetational characteristics. In spite of the texture diversity throughout their growth cycle, sugarcanes are nondescript among banana trees, cassava plants, rice and some other crops in mature period (Figure 2). Therefore, considering these factors, high extraction accuracy will be difficult to achieve if remote sensing image is the only source for deep learning.

According to statistical analysis, although sugarcane and certain crops have high degree of textural

similarity during a specified period, no crops are found consistent with sugarcane throughout the entire growth cycle in Guangxi. If temporal factor is brought into DCNN, its extraction accuracy would be significantly improved. In order to add temporal factor, grouping network method is adopted in the DCNN.
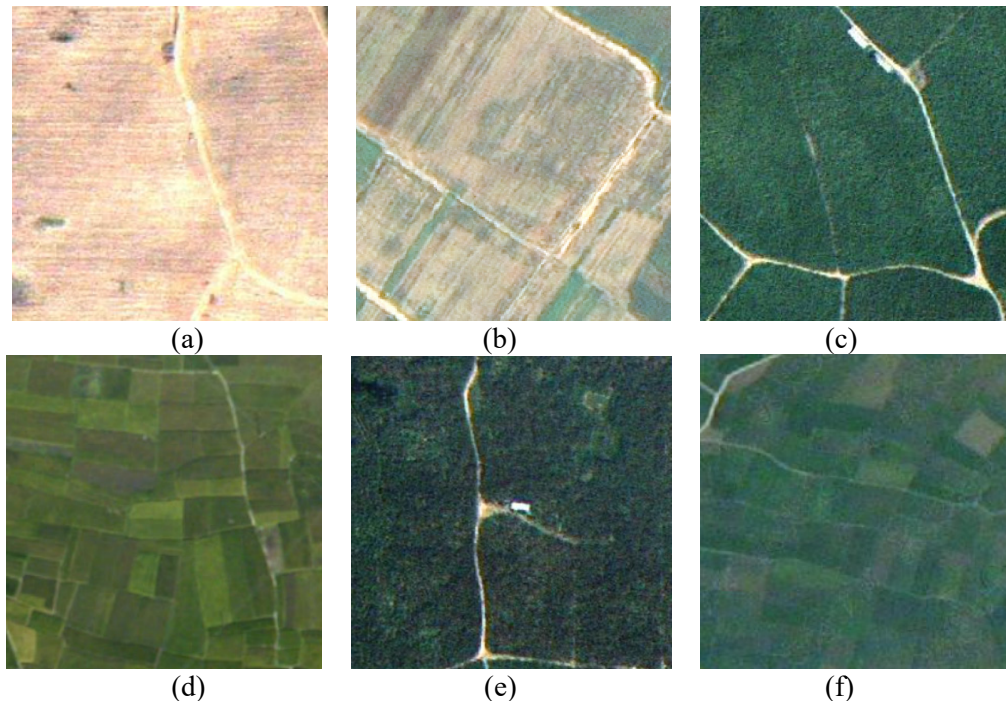

Figure 2. Sugarcane and Confused Vegetation Texture: (a) Texture of Sowing and seedling period, (b) Texture of During growing period, (c) Texture of Mature period, (d) Texture of rice, (e) Texture of Banana, (f) Texture of Cassava.

According to the periodic change of crop texture, the DCNN is designed as two parts. The first part is mainly used to extract hardened ground, buildings, bare lands, water and other non-vegetation areas, and then transfer the vegetated areas to the second part for further processing. The second part of the DCNN is dedicated to the temporal processing of vegetation. By grouping multi- period images and synthesizing the different temporal features of multi-stage images, the disturbance of confusing crops can be eliminated and the planting field of sugarcane can be extracted. The texture contrast between vegetation and non-vegetation is large after synthesizing many kinds of data because of the additional data such as near infrared, surface slope and so on. Through the two parts of the network, the regions with large contrast are segmented first, and then the sub-classes which are not easy to distinguish in vegetation are segmented, which can further improve the extraction accuracy. The structure of the neural network is shown in the Figure 3.
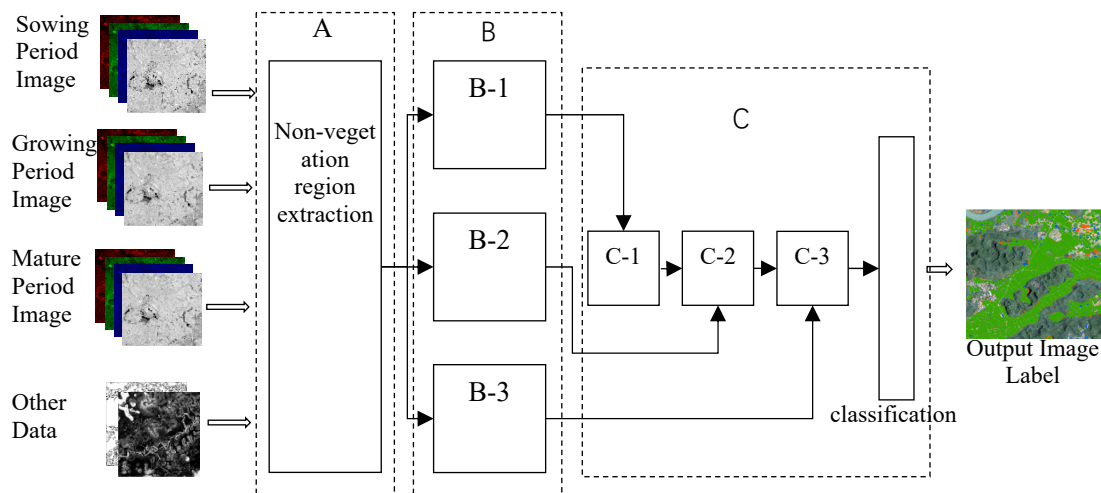
Figure 3. Deep Convolution Neural Network for Sugarcane Extraction

In this DCNN, the network of part A and part B is constructed based on the DeepLab V3+. Part A is mainly used to identify vegetation and non-vegetation areas. Image data are exported as vegetation and non-vegetation areas after group A processing. The network of part B is divided into three sub-networks: B-1, B-2 and B-3. Each sub-network receives images of sowing and seedling period, growing and maturing period for training and recognition respectively. Its purpose is to output sugarcane planting field in corresponding periods to part C. Tensor sequences of sugarcane field are constructed in Part C. The sequences are fused by 1*1 convolution. After 1*1 convolution, the final results are exported by classifier. In the process of training network, the first part and the second part of the DCNN are trained separately. The part B of the network is trained separately by three groups. The training samples of each group are selected as the requirements of the time period, and the ability of extracting features is strengthened by the separate training of the two parts.

3. Experiments and Analysis

3.1 Extraction results and analysis

In this study, all the trainings and experiments are implemented using Python 2.7 under the Windows server 2012 R2 environment. DCNN algorithm was implemented based on TensorFlow 1.8.0 and CUDA9.0. An Intel Xeon 6128/64 G/ Tesla M60 server was used for training and executing all the experiments. The experiment also compares sugarcane extraction performances of other methods, such as object-oriented and DeepLab V3+. The extraction results are shown in Figure 4.

The GF-2 and BJ-2 High-resolution remote sensing images of Wuming, Guangxi, which were capture in 2017 are select as experiment data. The experiment requires three periods of images in the same area, and the images should cover large sugarcane planting areas. Sixteen groups of experimental data are made according to the range of 5000 * 5000 pixels. Fourteen groups are randomly selected as training samples. The rest two groups are used for testify. 16 groups images of ground truth are made by visual interpretation and manual labeling according to the 2017 National Geography Condition Monitoring dataset and double-high sugarcane base boundary. In order to ensure sugarcane sample dataset's purity, all field roads, cloud covers and sporadic features should be excluded as much as possible during sample data production.
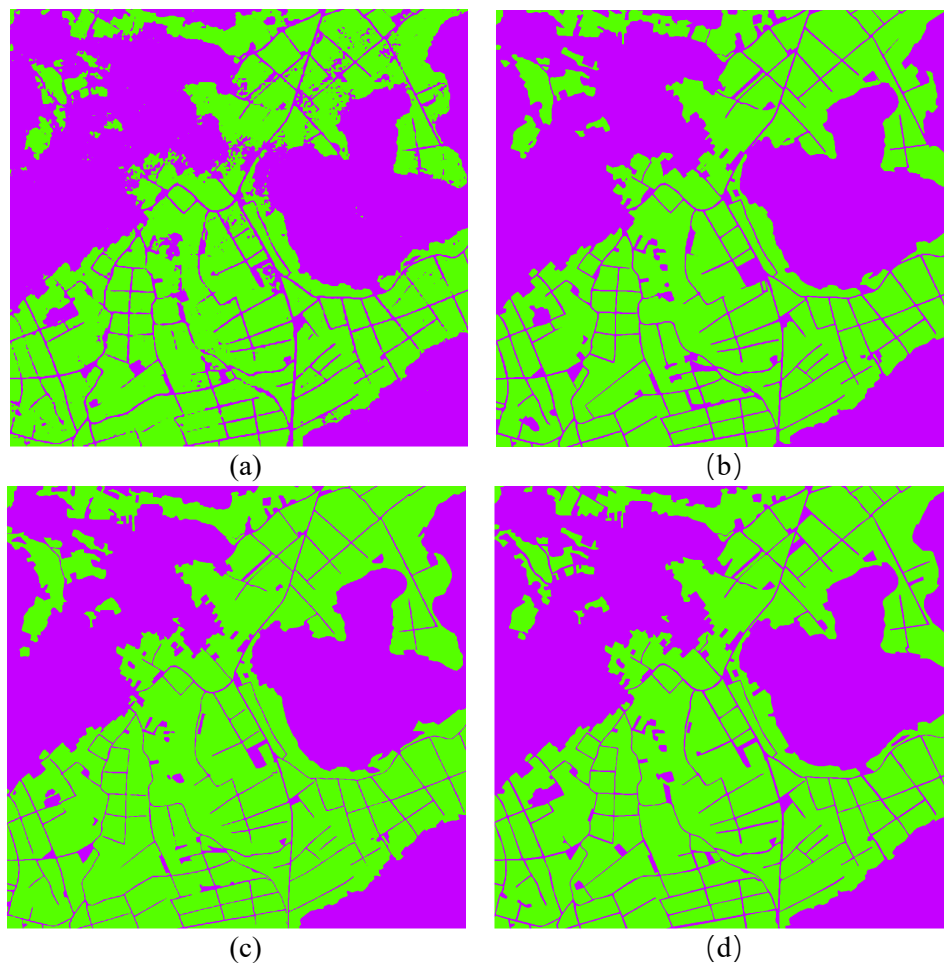
Figure 4. Comparison of Different Extraction Methods: (a)Object-oriented, (b) DeepLab V3+, (c) Our method, (d) Ground Truth.

Comparing to these methods, the proposed method has the best overall accuracy, reaching 94.32%, and the extracted fields are consistent with ground truth. Errors are mainly found in edge areas of sugarcane fields and small fields. No mistakes are found in large sugarcane field recognition. Both object-oriented method and DeepLab V3+ have classification errors in large fields, and the object-oriented method shows obvious zigzag shape on the edge. DeepLab V3+ has the fastest processing speed in terms of processing efficiency, followed by the proposed method. The reason for speed decrease is that the number of network parameter and layers is increased. Our DCNN improves accuracy of extraction with additional computation, resulting in a decrease in processing speed, yet still significantly faster than the object-oriented method. Table 1 shows the specific performance comparison results.

Table 1. Performance Comparison of Different Extraction Methods

Method	Accuracy	Recall	Processing time
Object-oriented	85.30%	85.77%	1356s
DeepLab V3+	90.94%	89.52%	112s
Our DCNN	94.32%	93.99%	289s

The experimental results show that the accuracy of the proposed method is improved effectively by adding temporal factors. From Table 2, we can find that the accuracy of our method is as the same as DeepLab V3+ when processing single period images, but the accuracy of our method is better than DeepLab V3+ when inputting three period images at the same time. Owing to that the feature extraction of DeepLab V3+ cannot process temporal factors, the extraction accuracy has not been improved. The object-oriented method cannot directly process multi-temporal data due to the

limitations of the method.

Table 2. Accuracy Comparison in Different Planting Periods

Data Period	Object-oriented	DeepLab V3+	Our DCNN
Ⅰ. Sowing and seedling	70.47%	78.76%	77.41%
Ⅱ. Growing	80.19%	84.46%	86.39%
Ⅲ. Mature	85.30%	90.94%	92.12%
Ⅰ+Ⅱ+Ⅲ	/	89.78%	94.32%

In terms of error distribution, the majorities are edge error, omission and misclassification（Figure 5）. Edge errors are defined when the output boundary exceeds 3 pixels from the edge of ground truth fields. Flowing the definition, the percentage of edge errors reaches 34% of the overall errors. It shows that although the introduction of multi-temporal improves extraction accuracy, the difference of sugarcane growth in different periods leads to edge confusion, which reduces extraction accuracy.
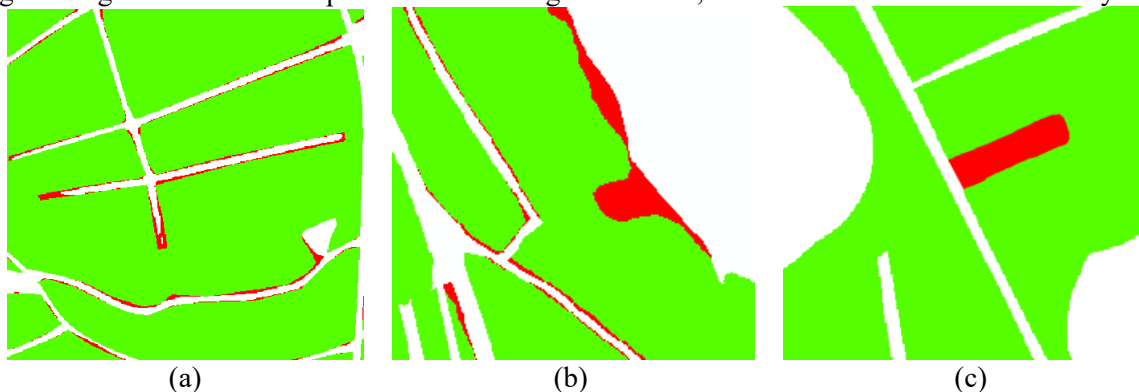

(a)   (b)   (c)

Figure 5. Major Error Type: (a) Edge error, (b) Omission, (c) Misclassification.

4.Conclusion

In this paper, a sugarcane planting field extraction method based on deep learning is proposed and temporal concept is introduced. Compared with the traditional object-oriented method, the extraction accuracy of sugarcane planting field is improved from 85.30% to 94.32% without considering the design features and extraction rules. The extraction accuracy is better than object-oriented method in the case of texture approximation. Comparing to deep learning method using DeepLab V3+ alone, the accuracy is improved from 90.94% to 94.32%. It shows that the accuracy can be further improved by adding temporal factors, which can be used in practical applications. Meanwhile, the difference of sugarcane growth in different periods leads to edge confusion, which reduces accuracy of edge extraction. In the next step, the edge confusion should be optimized to further improve the extraction accuracy.

References

[1] Li D.R., Tong Q.X., Li R.X., Gong J.Y., Zhang L.P. (2012) Current issues in high-resolution Earth observation technology. Sci China Earth Sci, 55: 1043-1051.
[2] PAGOT E., PESARESI M. (2008) Systematic Study of the Urban Post conflict Change Classification Performance Using Spectral and Structural Features in a Support Vector

Machine. IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing, 1(2): 120-128.

[3] Li D.R., Zhang L.P., Xia G.S. (2014) Automatic Analysis and Mining of Remote Sensing Big Data. Acta Geodaetica et Cartographica Sinica, 43(12):1211-1216.

[4] Lecun Y., Bengio Y., Hinton G. (2015) Deep learning. NATURE, 521(7553): 436-444.

[5] Girshick R., Donahue J., Darrell T., Malik J. (2014) Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus. pp. 580-587

[6] Gong J.Y., Ji S.P. (2018) Photogrammetry and Deep Learning, Acta Geodaetica et Cartographica Sinica, 47(06): 693-704.

[7] Gong J.Y., Ji S.P. (2017) From Photogrammetry to Computer Vision. Geomatics and Information Science of Wuhan University, 42(11): 1518-1522.

[8] Ball J.E., Anderson D.T,, Chan C.S. (2017) A Comprehensive survey of deep learning in remote sensing: theories, tools, and challenges for the community. Journal of Applied Remote Sensing, 11.

[9] Chen L., Papandreou G., Kokkinos I., Murphy K., Yuille A.L. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. IEEE Transactions on Pattern Analysis and Ma-chine Intelligence. 2018, 40(4): 834-848.

[10] Chen L.C., Papandreou G., Schroff F., Adam H. (2017) Rethinking Atrous Convolution for Semantic Image Seg-mentation. https://arxiv.org/abs/1706.05587.

[11] Chen L.C., Zhu Y., Papandreou G., et al. Encoder-Decoder with Atrous Separable Convolution for Se-mantic Image Segmentation. https://arxiv.org/abs/1802.02611.